

We put the SQL back in NoSQL*

@ApachePhoenix

http://phoenix.apache.org/

James Taylor (@JamesPlusPlus)

* Anybody else notice that all the NoSQL stores have SQL?

About James

- Architect at Salesforce.com
 - Part of the Big Data group
 - Lead of the Apache Phoenix project









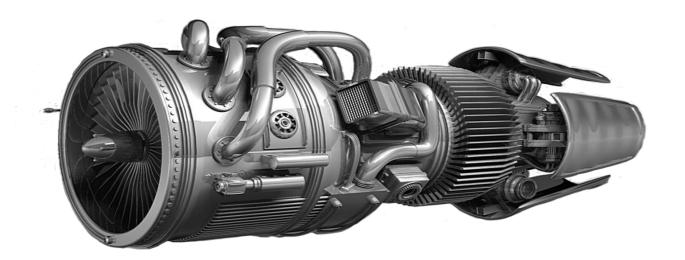
What is Apache Phoenix?

A relational database layer for Apache HBase



What is HEASE?

- High performance horizontally scalable byte store
- Suitable as store of record for mission critical data





What is Apache Phoenix?



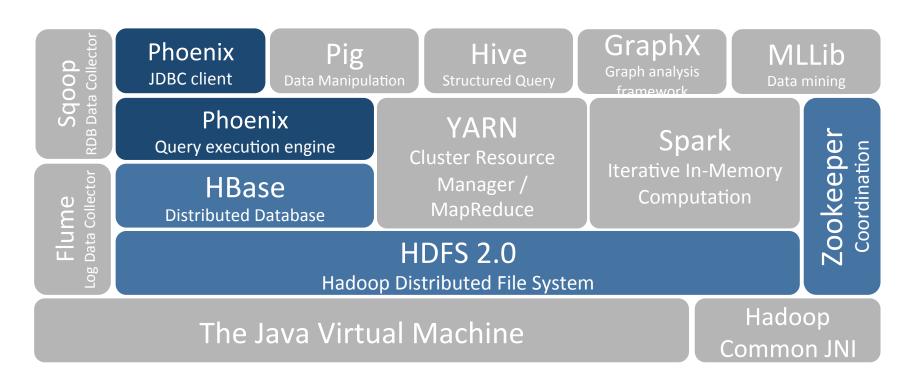


What is Apache Phoenix?

- A relational database layer for Apache HBase
 - Query engine
 - Transforms SQL into native HBase API calls
 - Pushes work to cluster for parallel execution
 - Metadata repository
 - Typed access to data stored in HBase tables
 - Support multi-tenancy modeled as SQL views
 - A JDBC driver
- A top level Apache Software Foundation project
 - Originally developed at Salesforce
 - Now a top-level project at the ASF
 - A growing community with momentum



Where Does Phoenix Fit In?



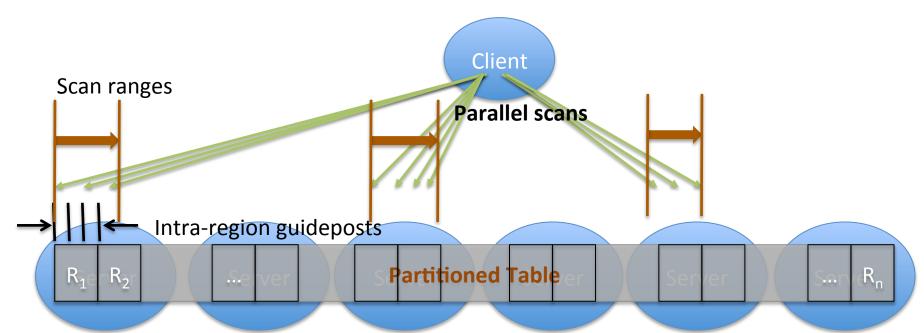


Why is Phoenix fast?

- Pushes down computation to region servers
 - Start/stop key range(s)
 - Time range min/max
 - Predicates
 - Aggregation
 - Sort
 - Limit
 - TopN
- Parallelizes query from client
 - Intra-region through statistics collection
- Supports secondary indexes
 - Global & co-located



Why is Phoenix fast?

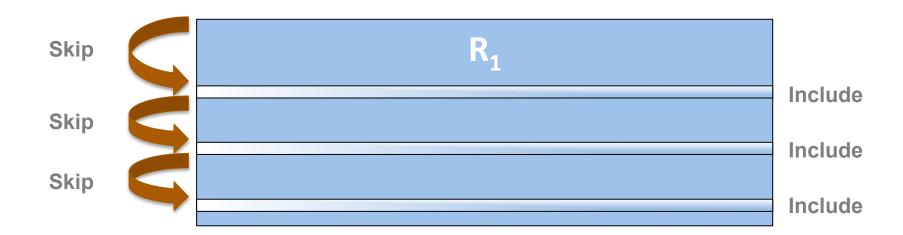


- Filter
- Aggregate
- Sort
- Limit



Skip Scans

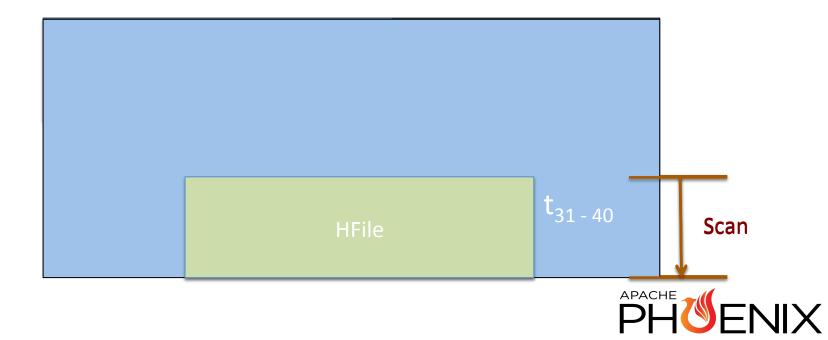
- Push key range sets through filter
- Use SEEK_NEXT_HINT to skip data





Filters

- Pushes down WHERE clause to server
- Filters entire HFile based on time-range



What's Next?

Big Data Landscape (Version 2.0) Analytics Applications -Infrastructure Analytics Solutions \ Data Visualization Ad Optimization 10gen DATASTAN DOSNO CIDUCET HADAPT
COUCHBASE COUDANT infochimps Q Palantir platfora Quid visual.ly aggregate Obluekai rocketfue Hortonworks centrifuge metalayer **Zettaset** A LATTICE Quantum4D Microsoft VISUAL Sair SCIENCE Yieldex | bloomreach p(k) Prior Knowledge DEVOLUTION MATLANT | bit.ly MPP Databases Management / Cluster Services Dataminr ndustry Applications KNEWTON Zest cash" OUTER THOUGHT I HPCC Systems Sentiment Analysis I **Analytics Services** Acunu numberFire MileSense puri THINK BIG GENERAL SENTIMENT PAR ACCEL. Climate Solutions Bloomberg GUARD **⊙**Stormpath accenture OPERA Application Service Providers 1 @ IMPERVA N NETEZZA I (i) collective[] DATADOG InfiniDB SOL Server Recorded Future **Data Sources** Place [C RADIUS / splunk > sumolog I DATAGUISE Real-Time (Crowdsourced SMB Analy panasas | CrowdFlower | Collection | G CONTINULTY II Analytics sum/ factual. knoema GNIP a nimblestorage | | DataKind DataMarket infochimps RJMe I aspera I mechanicalturk / I nodeable feedzai / kaggle Windows Azure Cross Infrastructure / Analytics SAP SSAS. IEM METAMARK ORACLE Microsoft vmware amazon JAWBONE RunKeeper + + fitt Open Source rojects Query / Data / Data Access | mongo

You are here



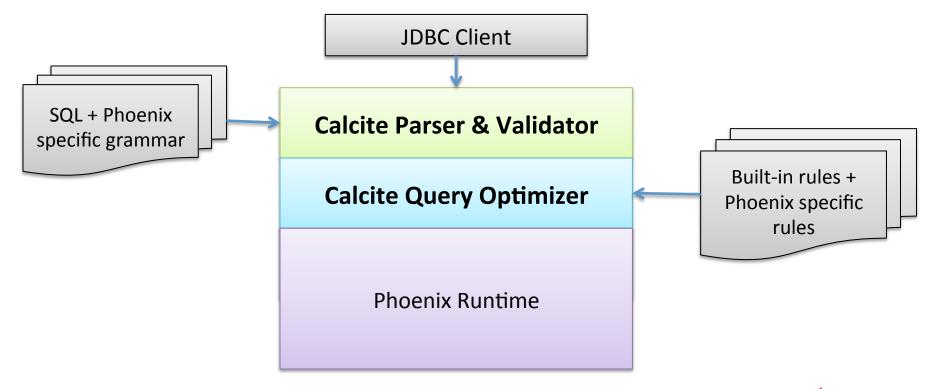
© Matt Turck (@mattturck) and ShivonZilis (@shivonz) Bloomberg Ventures

Introducing Apache Calcite

- Query parser, compiler, and planner framework
 - SQL-92 compliant
- Pluggable cost-based optimizer framework
 - Sane way to model push down through rules
- Interop with other Calcite adaptors
 - Already used by Drill, Hive, Kylin, Samza
 - Supports any JDBC source (RDBMS remember them ☺)
 - One cost-model to rule them all

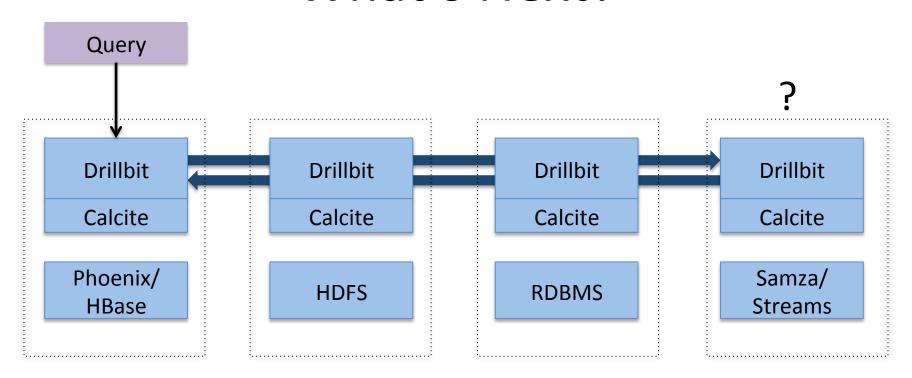


How does Phoenix plug in?





What's Next?





Thank you! Questions?



































